

УДК 517.15

Сидор В. –ст. гр. СНм-61

Тернопільський національний технічний університет імені Івана Пулюя

АНАЛІЗ ОСОБЛИВОСТЕЙ КЛАСИФІКАТОРІВ ТЕКСТУ

Науковий керівник: к.т.н., доцент Фриз М. Є.

Sydor V.

Ternopil Ivan Puluj National Technical University

ANALYSIS OF THE FEATURES OF CLASSIFIERS OF THE TEXT

Supervisor: associate professor Fryz M. Ye.

Ключові слова: Класифікація, Текст

Keywords: Classification, Text

У доповіді буде розглянуто класифікацію та види класифікації. Класифікація тексту - це процес призначення тегів або категорій тексту відповідно до його змісту. Класифікацію тексту можна зробити двома різними способами: ручною і автоматичною. Перший спосіб – анотатор, людина, яка інтерпретує зміст тексту і відповідно класифікує його. Переваги - може забезпечити хороші результати. Недоліки - займає багато часу і дорого обходиться. Другий спосіб - застосовує машинне навчання, обробку природної мови та інші методи автоматичного класифікації тексту. Переваги - набагато швидший та економніший.

Існує багато підходів до автоматичної класифікації текстів, які можна згрупувати в три різні типи систем: системи на основі правил, системи машинного навчання та гібридні системи.

Класифікатори не можуть працювати автоматично, вони повинні бути навчені, щоб мати можливість робити конкретні прогнози для текстів. Тренування класифікатора полягає у визначенні набору тегів, з якими буде працювати модель та створення асоціацій між фрагментами тексту і відповідними тегами;

Як тільки достатньо текстів було позначено, класифікатор може вчитися у цих асоціацій і починати робити прогнози з новими текстами.

Класифікація тексту може бути застосована до широкого кола завдань. У деяких випадках класифікатори працюють за лаштунками, щоб розширити можливості продукту чи програми.

Класифікатор тексту нічого не вартий без точних даних для навчання. Розповідаючи алгоритму, що ви очікуєте певного набору тегів як вихідних даних для певного тексту, він може навчитися розпізнавати шаблони в тексті, наприклад, почуття, яке виражає твіт або тематику тексту. Велику роль у точності класифікатора відіграє отримання правильних даних для навчання. Саме тому для покращення результатів роботи класифікатора потрібно використовувати правильний набір даних.

Література

1. Олексіїв А. А. Класифікація текстових документів на основі технології text mining / А. А. Олексіїв., 2016. – 114 с.
2. Text classification [Електронний ресурс] – Режим доступу до ресурсу: <https://monkeylearn.com/text-classification/>.